medRxiv preprint doi: https://doi.org/10.1101/2023.08.31.23294918; this version posted September 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

RESEARCH

Benefit-Risk Assessment of Medical Products Using Bayesian Multi-Criteria Augmented Decision Analysis (MCADA) for Clinical Development

Heather J Berringer[1,2] | **Ofir Harari**[1] | **Vinusha Kalatharan**[1] | **Awa Diop**[1] | **Edward J Mills**[3,4] | **Jay JH Park**[1,3]

[1]Core Clinical Sciences, Vancouver, BC, Canada

[2]Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada

[3]Department of Health Research Methodology, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

[4]Purpose Life Sciences, Vancouver, BC, Canada

Correspondence
Jay JH Park
Email: jay@coreclinicalsciences.com

Present address
400 - 601 West Broadway Vancouver, BC V5Z 4C2, Canada

Abstract

Multi-criteria decision analysis is a benefit-risk assessment tool that evaluates multiple competing benefit and risk criteria simultaneously. This has the potential to aid sponsors in making effective and informed go/no-go decisions for their clinical development program. This method involves assigning weights to various benefit and risk criteria based on their relative importance (utility weight) and summing them to compute a single utility score that represents the overall benefit-risk profile of the treatment. However, this approach is constrained to binary and continuous parameters. In this paper, we introduce a novel framework known as Bayesian Multi-Criteria Augmented Decision Analysis (MCADA), which extends existing methods to encompass time-to-event and ordinal outcomes while incorporating linear and novel non-linear functions in utility aggregation. This paper provides a comprehensive description of the statistical methodology behind the MCADA framework and demonstrates its application using IPD and aggregate data from two clinical trials. Our two case studies show that the MCADA framework can be effectively used to produce a single utility score that reflects the overall benefit-risk profile of the treatment using both IPD and aggregate data from trials. MCADA broadens the horizon of the existing MCDA framework by accommodating a wider range of data types and utility functions in the utility aggregation process.

KEYWORDS
Bayesian analysis, benefit-risk, decision making, multi-criteria decision analysis

1 | INTRODUCTION

At the end of each early phase clinical trials, the sponsors are positioned to decide whether to pursue a clinical trial of subsequent phase (go/no-go decision) that requires considerably larger resources[1]. However, making an optimal go/no-go decision can be limited by how clinical trials are often designed and analyzed. For instance, while information on several benefit and risk parameters are often collected and analyzed independently, most clinical trials are optimized based on a single parameter, with trials being powered based on a single primary outcome that acts as the driver for decision making.

Optimal go/no-go decision-making requires consideration of multiple benefit and risk criteria simultaneously[2]. Multi-criteria decision analysis (MCDA) has been proposed in the past as a benefit-risk assessment tool that evaluates multiple competing benefit and risk options concurrently[3–5]. The MCDA approach entails weighting multiple benefit and risk criteria based on their relative importance (utility weight) and adding them to a calculate a single utility score that reflects the treatment's overall benefit-risk profile. This quantitative framework has the potential to enhance internal decision-making for sponsors during their clinical development programs. Since 2011, the European Medicines Agency (EMA) have supplemented their decision-making process with MCDA methods, particularly when multitude of factors need to be taken into account to arrive at a decision[6]. Similarly, the FDA uses MCDA when facing challenging decisions and use it to achieve a more robust decision[6].

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.08.31.23294918; this version posted September 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Several MCDA methods have been proposed in the past to assess the benefit-risk profile of medical products[7–13]. However, currently available MCDA methodologies have limited applicability. Some common trial endpoints used in oncology include time-to-event outcomes, such as overall survival, and ordinal outcomes, such as the RECIST criteria. However, the application of MCDA is currently limited to binary and continuous parameters and often apply a linear utility function in the aggregation of the utility. Dichotomization of outcomes is often problematic because it results in lost information and reduced statistical power to detect a treatment effect[14]. Additionally, making an linearity assumption while aggregating treatment's benefits and risks can be unrealistic. For example, making the linearity assumption for an adverse event would imply that the disutility associated with the adverse event would increase constantly. Such implicit assumption on the preferences of decision-makers likely does not hold for all medical products. Experiencing adverse events up to a certain rate could be afforded since taking the drug could result in clinical improvement. Non-linear utility functions have been proposed[7,8,15], but suffer from unclear clinical interpretation. These limitations may have narrowed the applications and versatility of existing quantitative MCDA methods.

There is a need to expand the quantitative MCDA methods for medical products to capture the benefit-risk profiles of other outcomes beyond binary and continuous outcomes with linear and non-linear utility functions. In this paper, we propose a new framework called Bayesian Multi-Criteria Augmented Decision Analysis (MCADA) that extends the existing methods. In particular, MCADA follows the probabilistic MCDA approach of[16] and[7], that utilizes Bayesian modelling to estimate the probabilistic distributions of binary or continuous criteria variables. It then combines criterion subjective relative importance in the stakeholder eye with these distributions to generate the final utility score distribution for the treatment. Our MCADA framework can incorporate a broader range of endpoints, including time-to-event and ordinal outcomes, with both linear and interpretable non-linear utility functions. In this MCADA framework that we have developed, both individual patient-level data (IPD) and aggregate-level summary statistics reported from clinical trials can be incorporated, making MCADA an effective quantitative benefit-risk assessment tool for clinical development programs.

The objective of this paper is to describe the statistical methodology of the MCADA framework and to show its application using both IPD and aggregate data. The paper is organized as follows. In section 1, the statistical methodology of the MCADA framework is described. In section 2, a benefit-risk assessment in lung cancer treatment as a case study is provided. Another case study of MCADA, based on pancreatic cancer data, is presented in section 3. Finally, in section 4, a discussion and concluding remarks is provided.

2 | METHODS

2.1 | Overview of MCADA Framework

The MCADA framework is a quantitative benefit-risk assessment tool that help decision-makers evaluate alternatives based on multiple parameters simultaneously, rather than a single parameter. It extends the MCDA methodology to allow evaluation of outcomes beyond binary and continuous outcomes, and to allow statisticians to model a non-linear utility function.

The MCADA framework consists of the following: 1) Identification of all relevant benefit and risk parameters. 2) Ranking of the different parameters based on clinical importance to clinicians and patients. 3) Assigning weights to each parameter based on the ranking from step 2. 4) Deriving score distributions for each parameter using Bayesian analysis. 5) Normalizing the distributions derived in Step 4 to values between 0 and 1. 6) Calculating utility scores distributions by combining the weights and normalized distributions.

In the following sections, we will provide the technical details of MCADA.

2.2 | Binary, Continuous, and Ordinal Models

Given m treatments ($i = 1, \dots, m$) assessed via n different criteria ($j = 1, \dots, n$), the performance of treatment i on criteria j is denoted by ξ_{ij}. MCADA treats ξ_{ij} as a random variable and uses Bayesian modelling to assign a probability distribution to each ξ_{ij}, depending on the data type of the criteria. If criteria j produces binary data, ξ_{ij} will follow a beta distribution arising from a beta-binomial model with parameters determined by the number of events and patients per treatment, $\xi_{ij} \sim \textit{Beta}(\alpha_{ij}, \beta_{ij})$. If it is continuous, ξ_{ij} will follow a Student-t distribution informed by the mean, standard deviation, and number of patients assessed by criteria j. If j is an ordinal variable, ξ_{ij} will follow a Dirichlet distribution arising from a Dirichlet-multinomial model. The Dirichlet parameters will be informed by both the prior and the data, where the data are summarized as a vector of counts per category. For each treatment i, we have a vector indicating how the treatment performed on each criterion: $\xi_{\boldsymbol{i}} = (\xi_{i1}, \dots, \xi_{in})$.

medRxiv preprint doi: https://doi.org/10.1101/2023.08.31.23294918; this version posted September 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

2.3 | Time-to-Event Model

If criteria j is a time-to-event variable, the posterior of ξ_{ij} will be computed by first fitting a parametric survival curve based on a dependent gamma prior, then by calculating the median survival time for each posterior draw of individual survival curves.

To compute the posterior survival function for time-to-event data, we use the approach of Castillo and Van der Pas[17]. This method uses a dependent gamma prior to inform the heights of a piecewise exponential prior. Given the time interval of the data, $[0, \tau]$, divide the interval into K intervals of equal size: I_k, $k = 1, ..., K$. The value of the hazard function λ over each sub-interval is assigned a piecewise constant Gamma(α, β) prior, with shape parameter α and rate parameter β. The prior on the hazard λ is

$$\lambda(t) = \sum_{k=1}^{K} \lambda_k \mathbb{1}\{t \in I_k\} \tag{1}$$

with

$$\lambda_1 \sim Gamma(\alpha_0, \beta_0) \text{ and} \tag{2}$$

$$\lambda_k \sim Gamma(\alpha, \alpha/\lambda_{k-1}) \text{ for } k = 2, ..., K. \tag{3}$$

The dependent structure of the gamma prior can be seen from the mean and variance for $k = 2, \dots, K$:

$$E(\lambda_k \mid \lambda_{k-1}, ..., \lambda_1) = \lambda_{k-1} \tag{4}$$

$$Var(\lambda_k \mid \lambda_{k-1}, ..., \lambda_1) = \left(\frac{\lambda_{k-1}}{\alpha}\right)^2 \tag{5}$$

2.4 | Partial Value Function

Once posterior samples have been drawn for each ξ_{ij}, the simulated posterior values are passed through montonically increasingly partial value functions, $u_j(\cdot)$, that maps them onto the $[0, 1]$ interval. That is, if ξ_j' denotes the most desirable value for criteria j and ξ_j'' the least desirable value, we have $u_j(\xi_j') = 1$ and $u_j(\xi_j'') = 0$. All other utility values are given by

$$u_j(\xi_{ij}) = \frac{\xi_{ij} - \xi_j''}{\xi_j' - \xi_j''}. \tag{6}$$

Here, $u_j(\xi_{ij}) > u_j(\xi_{hj})$ indicates that treatment i is preferred to treatment h for criteria j.

This linear partial value function, while a common choice in decision analysis, has been shown in some cases to recommend a treatment with high risk or low benefit[18]. To address these issues, we introduced two non-linear transformations of the partial value functions: the Emax function and the logistic function. The Emax function is a non-linear function used for estimating dose-response curves[19]. In the context of MCADA, instead of measuring a dose-response relationship, we use the monotone, concave shape of Emax to measure a performance-utility relationship. Emax requires two parameter inputs, U_0 and U_{50}, where U_0 is the expected utility when the performance is lowest, and U_{50} is the performance that produces half of the maximum utility value. These parameters can be adjusted based on the clinical expectation of the performance-utility relationship in settings where this association is non-linear. The logistic function requires the same parameter inputs as Emax, with the addition of parameter δ, that determines the steepness of the logistic slope. The Emax function transformation of the partial value function is

$$\tilde{\xi}_i = U_0 + \frac{\xi_i}{U_{50} + \xi_i} \tag{7}$$

and the logistic function is

$$\tilde{\xi}_i = U_0 + \frac{U_{50}}{1 + \exp\left((U_{50} - \xi_i)/\delta\right)} \tag{8}$$

The linear and non-linear partial value functions used in MCADA can be seen in Figure 1.

2.5 | Linear Utility Aggregation

Using the vector $w = (w_1, \dots, w_n)$ provided by the stakeholder to assign weights to the respective criteria, a linear utility score $u(\xi_i, \boldsymbol{w})$ is then calculated for treatment i as a weighted average

medRxiv preprint doi: https://doi.org/10.1101/2023.08.31.23294918; this version posted September 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

$$u(\xi_i, \boldsymbol{w}) = \sum_{j=1}^{n} w_j u_j(\xi_{ij}) \tag{9}$$

Under the Bayesian model, the utility score is in fact a random variable, with distribution obtained from repeated sampling from the posteriors of the ξ_{ij}. For non-linear transformations, simply replace ξ_i with $\tilde{\xi}_i$ in the above equation. The utility scores for different treatments are compared graphically and through descriptive statistics.

3 | CASE STUDIES

In this section, the use of MCADA methods to support decision-making for clinical development programs in oncology using aggregate and IPD data of two phase II trials obtained using Project Data Sphere is illustrated.

Two benefit parameters of IPD data, overall survival (time-to-event outcome) and RECIST criteria (ordinal outcome), as well as one aggregate data risk parameter, grade 3 or higher adverse event rates (binary outcome) was considered for both case studies. Analytical hierarchy process (AHP) was used to derive the utility (relative importance) values of these parameters[20]. For both of our case studies, treatment performance on overall survival was considered twice as important as the RECIST criteria, and the tumor response measured by RECIST criteria was assumed to be twice as important as minimizing adverse events. Utility values of 0.57, 0.29, and 0.14 were assumed for overall survival, RECIST criteria, and grade 3/4 adverse events, respectively. For sensitivity analysis, all parameters were assumed to be of equal importance and was assigned a utility value of 0.33.

For overall survival, the most desirable median survival time was defined to be 20 months, and the least desirable median survival time was defined as 0 months. For RECIST criteria, the best response recorded was considered and relative utility values of 0, 0.33, 0.67, and 1 was assigned to progressive disease, stable disease, partial response, and complete response, respectively. The most desirable value was 0% and the least desirable value was 50% for adverse event rate, so a utility value of 0 would be assigned if the adverse event rate was 50% and higher.

For both case studies, utility distributions derived using the linear utility function was presented. The resulting utility score distributions for treatment versus control were all computed using 100,000 simulations. The non-linear performance-utility relationship for overall survival and adverse event rates, as well as the linear relationship for RECIST response is visualized in Figure 2. For all non-linear results, the Emax function with parameters $U_0 = 0$ and $U_{50} = 0.3$ was used for overall survival, and the logistic function with parameters $U_0 = 0.25$, $U_{50} = 0.5$, and $\delta = 0.05$ was used for adverse event rates. The resulting utility distributions from the non-linear functions can be found in the Appendix.

3.1 | Case Study 1: Extensive-Disease Small Cell Lung Cancer

An open-label, randomized phase II trial was conducted to evaluate the clinical efficacy and safety of a novel selective peptide antagonist, LY2510924, in combination with standard of care (SOC; i.e. carboplatin/etoposide) compared to SOC alone in patients with extensive-disease small cell lung cancer was (NCT01439568)[21]. A total of 94 patients were randomized to receive the novel therapy in combination with standard-of-care (N=47) or standard-of-care alone (N=43). The median overall survival was 9.72 (95% CI: 6.6, 11.7) months for the treatment arm and 11.1 (95% CI: 8.3, 13.4) months for the control arm. Overall response rate was 86.0% for the treatment arm (N=35/47) and 81.0% (N=34/42) for the control arm. The grade 3 or 4 adverse event occurred in 51% (N=24/47) in the treatment arm, and 30.2% (N=13/43) in the control arm.

The calculated utility score for the experimental treatment arm was lower (0.47; 95% credible interval [CrI]: 0.41, 0.53) than the control arm (0.56; 95% CrI: 0.47, 0.69) (Table 1). The results of the MCADA analysis showed that the novel therapy added no clinical utility when compared to the standard-of-care (Figure 3). Similarly, when using the non-linear functions for overall survival and adverse events, the mean utility score for the treatment arm was lower (0.47; 95% CrI: 0.39, 0.58) than the control arm (0.60; 95% CrI: 0.44, 0.69) (Figure A2 in Appendix).

For LY+SOC, the mean (95% CrI) utility scores was 0.50 (0.40 0.60) for overall survival, 0.65 (0.62, 0.68) for tumor response, and 0.05 (0.00, 0.26) for adverse events. For SOC, the mean (95% CrI) was 0.56 (0.40, 0.75) for overall survival, 0.62 (0.56, 0.67) for tumor response, and 0.39 (0.12, 0.64) for adverse events (Figure 3). When computing the utility scores using non-linear functions, the Emax function was used for overall survival and the logistic function was used for adverse events. For the treatment arm, the mean (95% CrI) utility score was 0.50 (0.31, 0.69) for overall survival and 0.10 (0.00, 0.46) for adverse events. For the control arm, the mean (95% CrI) was 0.60 (0.44, 0.69) for overall survival and 0.54 (0.28, 0.68) for adverse events. Comparing the utility score between treatment and control arm of each component independently when using both linear and non-linear functions reiterates that SOC had a superior benefit-risk profile compared to LY+SOC (Figure 3).

The difference in utility scores between LY+SOC and SOC was computed, and the mean (95% CrI) difference in utility was -0.09 (-0.22, 0.02). The utility score distributions were also computed using equal weighting for each parameter, as a sensitivity

medRxiv preprint doi: https://doi.org/10.1101/2023.08.31.23294918; this version posted September 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

analysis (see Appendix). Results of the sensitivity analysis showed that the mean (95% CrI) was 0.38 (0.34, 0.49) for the LY+SOC and 0.52 (0.41, 0.63) for the SOC arm reiterating SOC has superior benefit-risk profile than LY+SOC.

3.2 | Case Study 2: Metastatic Pancreatic Ductal Adenocarcinoma

A randomized, multicenter phase II trial was conducted to evaluate the safety and efficacy of CO-101, a lipid-drug conjugate, compared to gemcitabine among patients with untreated metastatic pancreatic ductal adenocarcinoma and lower hENT1 expression (NCT01124786). A total of 367 patients were randomized at an equal ratio (1:1) to CO-101 (treatment) versus gemcitabine (control). This trial was powered for the overall survival outcome. Enrolling 360 patients would result in 90% power at 5% two-sided type I error rate, if the median overall survival was 7.7 months for the treatment arm and 6.0 months for the control arm.

The median overall survival (OS) was longer for the treatment arm compared to the control arm but the difference was not statistically different (hazard ratio [HR] [95% CI]: 1.072 (0.86, 1.34); median [95% CI] 5.2 (4.5, 6.2) months for treatment arm and 6.0 (5.2, 6.8) months for control arm). Overall response rate was 16.2% (N=25/154) for the treatment arm and 28.8% (N=41/142) for the control arm. The grade 3 or 4 adverse event occurred in 45.8% (N=82/179) in the treatment arm, and 45.3% (N=82/181) in the control arm.

The overall utility score that accounts for overall survival, tumor response, and adverse events for CO-101 versus gemcitabine is shown in Figure 4D. Although the primary outcome of overall survival alone did not provide a clear indication of whether the treatment or control arm is superior, the benefit-risk profile of CO-101 compared to control became more evident when examining the MCADA results. While the differentiation between the treatment and control arm was not clear based on the primary outcome of overall survival alone, the benefit-risk profile of CO-101 is clearer with the MCADA results. The CO-101 had a much lower mean overall utility score [95% CrI] (0.43 (0.38, 0.50)) than the gemcitabine (control) arm (0.50 (0.43, 0.58)), indicating that CO-101 did not show any additional clinical utility compared to control (Figure 4).

Comparing the individual criterion performances again demonstrated the superiority of the control arm compared to the treatment arm. For CO-101, the mean (95% CrI) utility score was 0.61 (0.50, 0.70) for overall survival, 0.27 (0.23, 0.30) for RECIST response, and 0.09 (0.00, 0.23) for adverse event. For gemcitabine, the mean (95% CrI) was 0.67 (0.60, 0.80) for overall survival, 0.34 (0.31, 0.38) for RECIST response, and 0.10 (0.00, 0.24) for adverse event (Figure 4).

The difference in utility scores between the treatment and control arms were also computed, and the mean (95% CrI) difference in utility was -0.06 (-0.15, 0.04). All utility score computations showed that gemcitabine had a superior benefit-risk profile compared to CO-101.

4 | DISCUSSION

Clinical trials are often powered on the primary outcome that plays a central role in help deciding whether to proceed further or to halt the clinical development program. However, there are scenarios where the results of the primary outcome in clinical trials are not definitely conclusive, making the decision-making process challenging. This paper introduces the MCADA method, which addresses this challenge by simultaneously evaluating multiple benefit and risk criteria, aiding in more informed decision-making. This paper provides a comprehensive description of the innovative MCADA method and demonstrates its practical application using both IPD and aggregate data from two clinical trials.

Both case studies presented in the data demonstrate the successful application of MCADA across various data types, encompassing both IPD and aggregated data, while accommodating both linear and non-linear utility functions. The case studies also illustrates that the proposed MCADA methods is particularly valuable in making a more decisive judgement when the results of the trial's primary outcome are inconclusive. More specifically, in a phase II clinical trial comparing the safety and efficacy of LY2510924 and SOC combination therapy with SOC alone, the primary outcome of overall survival indicated a shorter duration in the treatment arm, though this difference lacked statistical significance. Relying solely on the trial's primary outcome data didn't provide a clear basis for deciding whether to continue or halt the trail. However, MCADA unequivocally demonstrated that not only did the overall utility score favor the control arm, but the utility scores for each component consistently indicated a superior benefit-risk profile for the control arm compared to the treatment arm. The same trend was observed for the second case study.

The main strength of the innovative MCADA methodology compared to currently existing benefit risk assessment tools, such as MCDA, is in its capacity to offer clinical utility across a wider spectrum of data types and use both linear and non-linear functions to model utility. It can also effectively use both IPD and aggregated data. More specifically, the MCADA method expands upon MCDA to include ordinal and time-to-event variables; this is a vital improvement to existing MCDA frameworks[7,8,16], as many studies use time-to-event or ordinal endpoints depending on the trial phase. By including a more diverse array of data type in the benefit-risk assessment, MCADA can guide decision-making for a wider range of medical products, particularly in the face of progressively intricate scenarios[6,22]. Furthermore, the addition of the Emax and logistic transformation functions allows

medRxiv preprint doi: https://doi.org/10.1101/2023.08.31.23294918; this version posted September 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

for non-linear modelling of the performance-utility relationship that is easily adjusted for each endpoint. These functions have a straightforward interpretation, whereas other non-linear utility functions suffer from non-intuitive interpretation of the relationship between the treatment performances and the utility values[15]. Finally, when adjusting the clinical preference (utility weights), the MCADA results showed sensitivity to different weighting schemes, which enables greater control of the clinical relevance of the criterion[23]. For instance, in the first case study, the treatment arm had significantly larger adverse events rate, but closer response rate and overall survival when compared to the SOC arm. When weighted equally, the adverse events utility scores had a greater influence on the utility scores, widening the difference in utility distributions between the arms. Given the costly nature of late-stage clinical trials, MCADA can potentially save considerable resources by ensuring sponsors make optimal decisions based on all available information earlier in the trial process. MCADA offers a comprehensive assessment approach for the efficient evaluation of medical products by employing a more holistic definition of success that considers treatment effect significance, clinical relevance, and favorable benefit-risk balance.

There are certain limitations to the MCADA method. First, robustness of MCADA remains to be tested using a simulation study. Second, eliciting relative importance of each criteria can be a difficult and subjective process. MCADA requires consideration of multiple sensitivity analyses with varying utility values for different criteria. To address some of the limitations, MCADA may benefit from adding a Scalar-Loss Score (SLoS) utility function[9]. Instead of using a linear utility score with the goal of finding a treatment with the highest utility, the SLoS method lends to finding a treatment with the lowest loss. It offers the same advantages as the linear utility function, but avoids recommendations of non-effective or unsafe treatments and uses a convex preferencing between safety and efficacy. In other words, it can tolerate larger increases in risk if it corresponds to an increase in benefit when the initial benefit is minimal. Additionally, other multi-criteria benefit-risk methods, such as the stochastic multi-criteria acceptability analysis (SMAA)[24] and Dirichlet SMAA[7] allow for partial or unknown criteria weights. The SMAA method considers the criteria weighting to be another source of uncertainty, and the utility score distributions are estimated for all possible weighting combinations[24,25]. However, this makes SMAA increasingly complex as more criteria are added, and can have a high degree of uncertainty in its results[16]. The Dirichlet SMAA is a generalization of MCDA and SMAA, where instead of eliciting criteria preferences from clinicians, it applies a Dirichlet distribution to the criteria weights. The parameters of the Dirichlet distribution are varied to produce a best-guess weight set, with the interpretation that the precision parameter of the Dirichlet distribution is the strength of confidence decision makers have in their criteria preferences[7]. Both the SMAA and Dirichlet SMAA could be included into MCADA to be used as sensitivity analysis, as the models become complex by accounting for several sources of uncertainty[15].

5 | CONCLUSION

The MCADA method is comprehensive tool to assess the benefit and risk profile of medical products. It considers various benefit and risk parameters concurrently, handles diverse data types, and can incorporate both linear and non-linear utility models. Furthermore, it doesn't solely rely on IPD but also accommodates aggregated data. This method proves invaluable in assisting sponsors in making informed go/no-go decisions throughout the clinical development phase. Its particular strength lies in its ability to provide clarity when the primary outcome results of the trial are inconclusive, aiding in the decision-making process regarding whether to proceed with the next steps or not.

AUTHOR CONTRIBUTIONS

JJHP had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: HJB and JJHP Acquisition, analysis, or interpretation of data: HJB, OH, VK, AD, EJM, and JJHP Drafting of the manuscript: HJB, OH, VK, AD, EJM, and JJHP Critical revision of the manuscript for important intellectual content: HJB, OH, VK, AD, EJM, and JJHP
Statistical analysis: HJB, OH, AD, and JJHP Obtained funding: JJHP. Administrative, technical, or material support: OH, JJHP Supervision: JJHP

FINANCIAL DISCLOSURE

None reported.

CONFLICT OF INTEREST

The authors declare having no conflict of interest.

medRxiv preprint doi: https://doi.org/10.1101/2023.08.31.23294918; this version posted September 2, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

REFERENCES

1. Chuang-Stein C, Kirby S, French J, Kowalski K, Marshall S, Smith MK, et al. A quantitative approach for making go/no-go decisions in drug development. Drug Information Journal. 2011;45(2):187–202.
2. Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value trade-offs. Cambridge university press; 1993.
3. Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of medicines–part 1: the development of a new model using multi-criteria decision analysis. Pharmacoepidemiology and drug safety. 2007;16(S1):S2–S15.
4. Kolasa K, Zah V, Kowalczyk M. How can multi criteria decision analysis support value assessment of pharmaceuticals?-findings from a systematic literature review. Expert Review of Pharmacoeconomics & Outcomes Research. 2018;18(4):379–391.
5. Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. Value in health. 2012;15(8):1172–1181.
6. Chisholm O, Sharry P, Phillips L. Multi-criteria decision analysis for benefit-risk analysis by national regulatory authorities. Frontiers in medicine. 2022;8:820335.
7. Saint-Hilary G, Cadour S, Robert V, Gasparini M. A simple way to unify multicriteria decision analysis (MCDA) and stochastic multicriteria acceptability analysis (SMAA) using a Dirichlet distribution in benefit–risk assessment. Biometrical Journal. 2017;59(3):567–578.
8. Saint-Hilary G, Robert V, Gasparini M. Decision-making in drug development using a composite definition of success. Pharmaceutical Statistics. 2018;17(5):555–569.
9. Saint-Hilary G, Robert V, Gasparini M, Jaki T, Mozgunov P. A novel measure of drug benefit–risk assessment based on scale loss score. Statistical methods in medical research. 2019;28(9):2738–2753.
10. Mussen F, Salek S, Walker S. Development and application of a benefit-risk assessment model based on multi-criteria decision analysis. Benefit-risk appraisal of medicines: a systematic approach to decision-making. 2008;p. 111–119.
11. Tervonen T, Van Valkenhoef G, Buskens E, Hillege HL, Postmus D. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. Statistics in Medicine. 2011;30(12):1419–1428.
12. Nixon R, Dierig C, Mt-Isa S, Stöckert I, Tong T, Kuhls S, et al. A case study using the PrOACT-URL and BRAT frameworks for structured benefit risk assessment. Biometrical Journal. 2016;58(1):8–27.
13. Marcelon L, Verstraeten T, Dominiak-Felden G, Simondon F. Quantitative benefit–risk assessment by MCDA of the quadrivalent HPV vaccine for preventing anal cancer in males. Expert Review of Vaccines. 2016;15(1):139–148.
14. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ. 2006;332(7549):1080.
15. Menzies T, Saint-Hilary G, Mozgunov P. A comparison of various aggregation functions in multi-criteria decision analysis for drug benefit–risk assessment. Statistical methods in medical research. 2022;31(5):899–916.
16. Waddingham E, Mt-Isa S, Nixon R, Ashby D. A Bayesian approach to probabilistic sensitivity analysis in structured benefit-risk assessment. Biometrical Journal. 2016;58(1):28–42.
17. Castillo I, van der Pas S. Multiscale Bayesian survival analysis. The Annals of Statistics. 2021;49(6):3559–3582.
18. Marsh KD, Sculpher M, Caro JJ, Tervonen T. The use of MCDA in HTA: great potential, but more effort needed. Value in Health. 2018;21(4):394–397.
19. Macdougall J. Analysis of dose–response studies—Emax model. In: Dose finding in drug development. Springer; 2006. p. 127–145.
20. Saaty RW. The analytic hierarchy process—what it is and how it is used. Mathematical modelling. 1987;9(3-5):161–176.
21. Salgia R, Stille JR, Weaver RW, McCleod M, Hamid O, Polzer J, et al. A randomized phase II study of LY2510924 and carboplatin/etoposide versus carboplatin/etoposide in extensive-disease small cell lung cancer. Lung Cancer. 2017;105:7–13.
22. Hughes D, Bayoumi A, Pirmohamed M. Current assessment of risk–benefit by regulators: is it time to introduce decision analyses? Clinical pharmacology & therapeutics. 2007;82(2):123–127.
23. Broekhuizen H, IJzerman MJ, Hauber AB, Groothuis-Oudshoorn CG. Weighing clinical evidence using patient preferences: an application of probabilistic multi-criteria decision analysis. PharmacoEconomics. 2017;35:259–269.
24. Lahdelma R, Hokkanen J, Salminen P. SMAA-stochastic multiobjective acceptability analysis. European Journal of Operational Research. 1998;106(1):137–143.
25. Tervonen T, Lahdelma R. Implementing stochastic multicriteria acceptability analysis. European Journal of Operational Research. 2007;178(2):500–513.

TABLE 1 Summary of statistics from extensive-disease small cell lung cancer trial.

Statistics	Treatment (n=47)	Control (n=43)
Overall Survival, median (95% CI) (in months)	9.72 (6.6, 11.7)	11.1 (8.3, 13.4)
RECIST Response, n (%)	35 (86)	34 (81)
Grade 3 or 4 Adverse Event, n (%)	24 (51)	13 (30)

Abbreviation: confidence interval, CI.

medRxiv preprint doi: https://doi.org/10.1101/2023.08.31.23294918; this version posted September 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

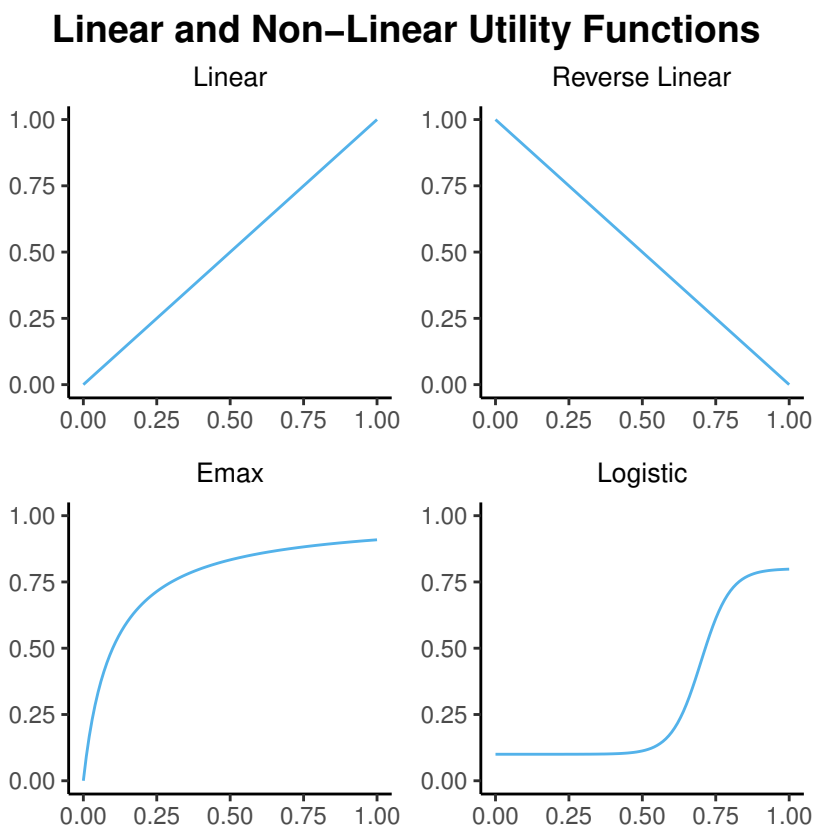

FIGURE 1 Linear and non-linear partial value function shapes. Parameter values to generate the Emax curve: $U_0 = 0$ and $U_{50} = 0.1$. Parameter values to generate this logistic curve: $U_0 = 0.1$, $U_{50} = 0.7$, and $\delta = 0.05$.

TABLE 2 Summary of utility score results from case study 1 using linear utility functions.

Criteria	Treatment Utility Score, mean (95% CrI)	Control Utility Score, mean (95% CrI)
Overall Survival	0.50 (0.40, 0.60)	0.56 (0.40, 0.75)
RECIST Response	0.60 (0.54, 0.65)	0.62 (0.56, 0.67)
Grade 3 or 4 Adverse Event	0.05 (0.00, 0.26)	0.39 (0.12, 0.64)
Overall Utility	0.47 (0.41, 0.53)	0.56 (0.47, 0.69)

Abbreviation: credible interval, CrI.

medRxiv preprint doi: https://doi.org/10.1101/2023.08.31.23294918; this version posted September 2, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

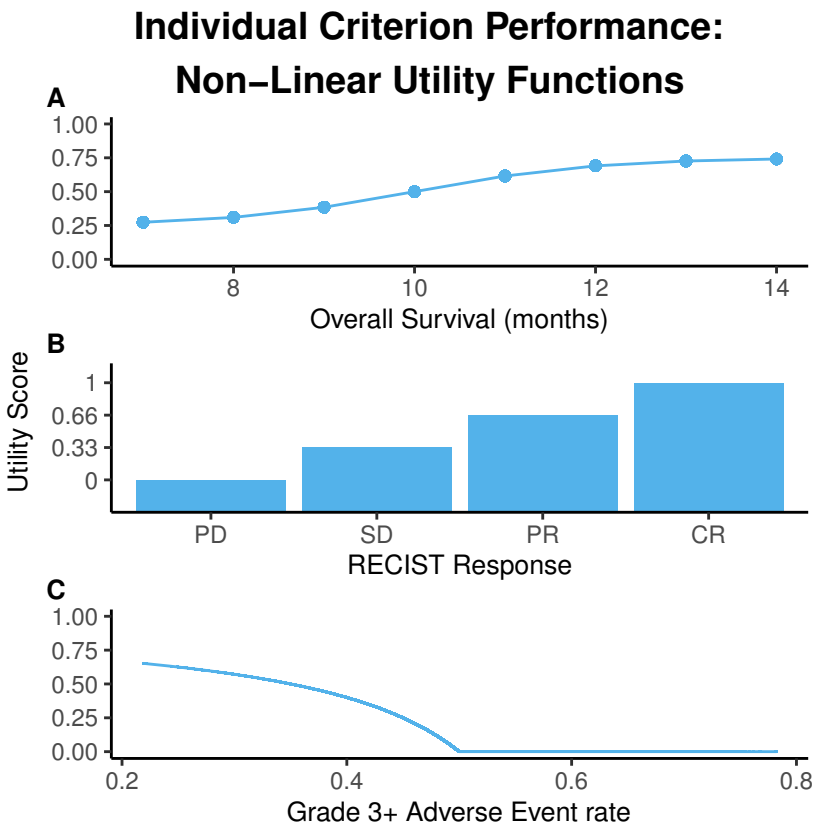

FIGURE 2 The performance-utility relationship for A) overall survival, B) RECIST response, and C) grade 3+ adverse events. A) results from a logistic transformation with $U_0 = 0.25$, $U_{50} = 0.5$, and $\delta = 0.05$, B) results from a linear function, and C) results from an Emax transformation with $U_0 = 0$ and $U_{50} = 0.3$.

medRxiv preprint doi: https://doi.org/10.1101/2023.08.31.23294918; this version posted September 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

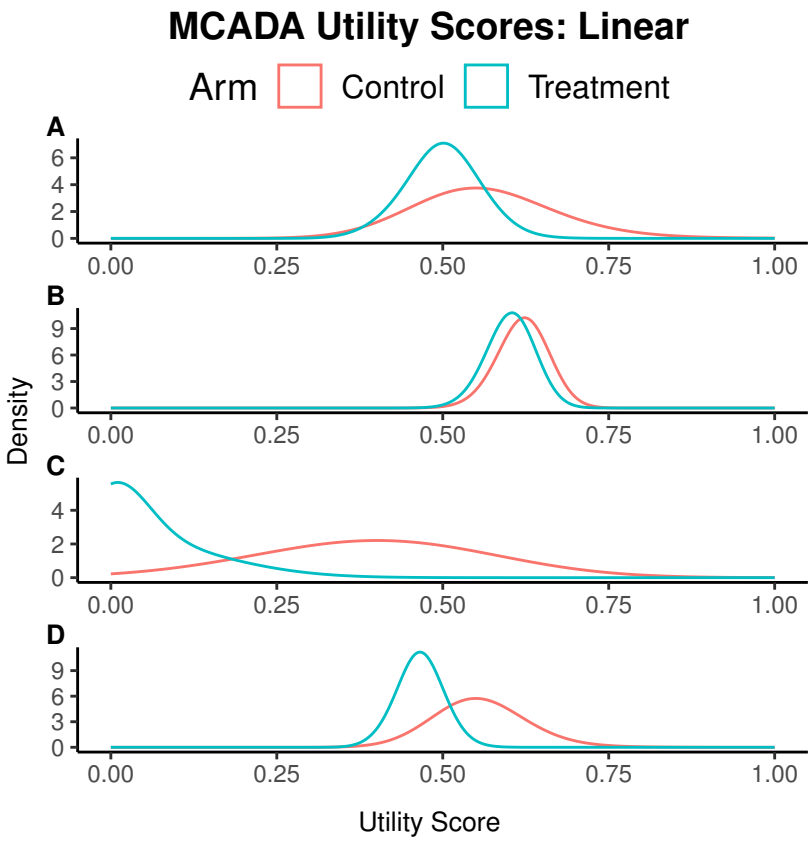

FIGURE 3 The utility score distributions of LY+SOC (treatment) and SOC (control) for (A) overall survival, (B) best tumor response measured by RECIST, (C) grade 3 or 4 adverse events, and (D) overall utility in patients with extensive disease non-small cell lung cancer. The utility values (relative importance) of 0.57, 0.29, and 0.14 were assumed for overall survival, best tumor response measured by RECIST, and grade 3 or 4 adverse event, respectively.

medRxiv preprint doi: https://doi.org/10.1101/2023.08.31.23294918; this version posted September 2, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

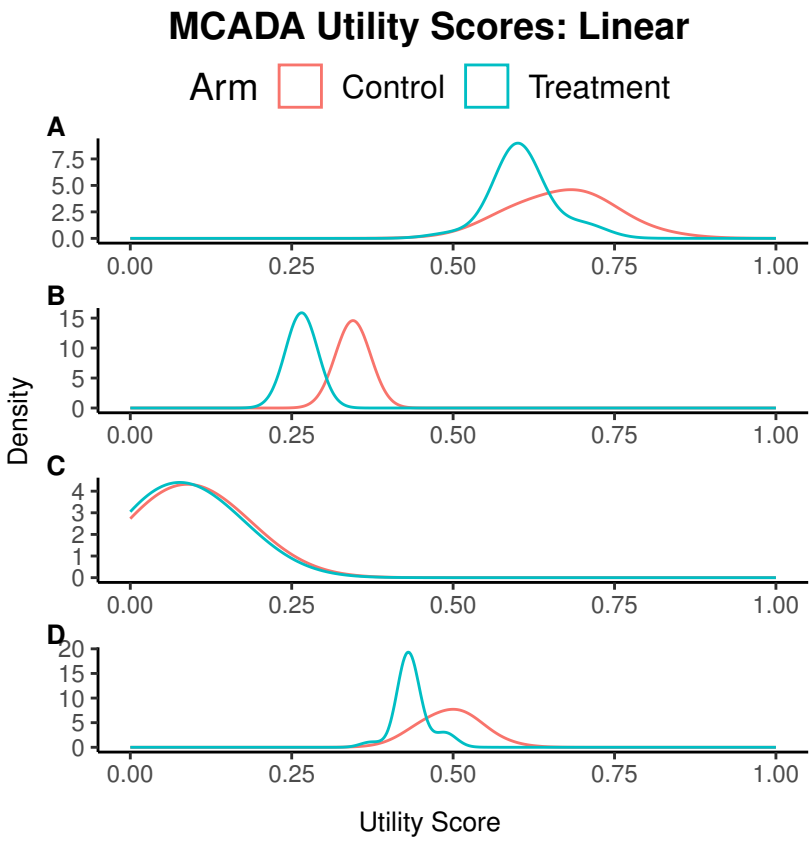

FIGURE 4 The utility score distributions of CO-101 (treatment) and gemcitabine (control) for (A) overall survival, (B) best tumor response measured by RECIST, (C) grade 3 or 4 adverse events, and (D) overall utility in patients with metastatic pancreatic ductal adenocarcinoma. Utility values (relative importance) of 0.57, 0.29, and 0.14 were assumed for overall survival, RECIST criteria, and grade 3 or 4 adverse events, respectively.